3A

(12) EUROPEAN PATENT APPLICATION

(72) Inventor: Barr, Philip J.
825 Lincoln Way
San Francisco California(US)

(72) Inventor: Merryweather, James P.
1427 Walnut Street
Berkeley California(US)

(72) Inventor: Mullenbach, Guy
1324 Grove Street
Berkeley California(US)

(72) Inventor: Urdea, Mickey S.
209 Irving Street
San Francisco California(US)

(74) Representative: Glawe, Delfs, Moll & Partner
Patentanwälte
Rothenbaumchaussee 58 Postfach 2570
D-2000 Hamburg 13(DE)

(54) Hybrid DNA synthesis of mature insulin-like growth factors.

(57) Methods and compositions are provided for efficient production of human insulin-like growth factor. Synthetic IGF I and IGF II genes are joined to leader and processing signals which provide for expression and secretion of the gene product in yeast. Enhanced yields of the product may then be recovered from the nutrient medium.

Yeast strains *S. cerevisiae* AB103 (pYIGF-I-10/1) and AB103 (pYIGF-II-10/1) were deposited at the American Type Culture Collection on *April 23, 1983* and granted Accession Nos. *20673* and *20674*, respectively.

EP 0 123 228 A2

describe the sequence encoding for the α-factor and spacers between two of such sequences.

SUMMARY OF THE INVENTION

Methods and compositions are provided for the
5 efficient production of mature human insulin-like growth factor (IGF). In particular, expression of a "pre"-IGF I and "pre"-IGF II in a yeast host facilitates secretion of the polypeptides into the nutrient medium. DNA constructs are generated by joining DNA sequences
10 from diverse sources, including both natural and synthetic sources. The resulting DNA constructs are stably replicated in the yeast and provide efficient, high level production of processed "pre"-polypeptides which may be isolated in high yield from the nutrient
15 medium.

DESCRIPTION OF THE SPECIFIC EMBODIMENTS

DNA sequences capable of expressing human insulin-like growth factors (IGF I and II) are provided. These DNA sequences can be incorporated into vectors,
20 and the resulting plasmids used to transform susceptible hosts. Transformation of a susceptible host with such recombinant plasmids results in expression of the insulin-like growth factor gene and production of the polypeptide product.
25 In particular, novel DNA constructs are provided for the production of the precursor polypeptides ("pre"-IGF I and "pre"-IGF II) in a yeast host capable of processing said precursor polypeptides and secreting the mature polypeptide product into the nutrient
30 medium. The DNA constructs include a replication system capable of stable maintenance in a yeast host, an efficient promoter, a structural gene including leader and processing signals in reading frame with said structural gene, and a transcriptional terminator
35 sequence downstream from the structural gene.

Optionally, other sequences can be provided for tran-
scriptional regulation, amplification of the gene,
exogenous regulation of transcription, and the like.
By "pre"-IGF I and "pre"-IGF II, it is meant that the

5   DNA sequence encoding for the mature polypeptide is
joined to and in reading frame with a leader sequence
including processing signals efficiently recognized by
the yeast host. Thus, "pre" denotes the inclusion of
secretion and processing signals associated with a

10 yeast host and not any processing signals associated
with the gene encoding the polypeptide of interest.

In preparing the DNA construct, it is necessary
to bring the individual sequences embodying the replica-
tion system, promoter, structural gene including leader

15 and processing signals, and terminator together in a
predetermined order to assure that they are able to
properly function in the resulting plasmid. As described
hereinafter, adaptor molecules may be employed to
assure the proper orientation and order of the sequences.

20 The IGF I and IGF II genes which are employed
may be chromosomal DNA, cDNA, synthetic DNA, or combina-
tions thereof. The leader and processing signals will
normally be derived from naturally occurring DNA se-
quences in yeast which provide for secretion of a poly-

25 peptide. Such polypeptides which are naturally secreted
by yeast include α-factor, a-factor, acid phosphatase
and the like. The remaining sequences which comprise
the construct including the replication system, promoter,
and terminator, are well known and described in the

30 literature.

Since the various DNA sequences which are
joined to form the DNA construct of the present invention
will be derived from diverse sources, it will be
convenient to join the sequences by means of connecting

35 or adaptor molecules. In particular, adaptors can be
advantageously employed to connect the 3'-end of the
coding strand of the leader and processing signal

sequence to the 5'-end of the IGF coding strand together
with their respective complementary DNA strands. The
leader and processing signal sequence may be internally
restricted near its 3'-terminus so that it lacks a

5   predetermined number of base pairs of the coding
region. An adaptor can then be constructed so that,
when joining the leader and processing sequence to the
IGF coding strand, the missing base pairs are provided
and the IGF coding strand is in the proper reading

10 frame relative to the leader sequence. The synthetic
IGF coding region and/or the adaptor at its 3'-end will
provide translational stop codons to assure that the
C-terminus of the polypeptide is the same as the
naturally occurring C-terminus.

15   The adaptors will have from about 5 to 40
bases, more usually from about 8 to 35 bases, in the
coding sequence and may have either cohesive or blunt
ends, with cohesive ends being preferred. Desirably,
the termini of the adaptor will have cohesive ends

20 associated with different restriction enzymes so that
the adaptor will selectively link two different DNA
sequences having the appropriate complementary cohesive
end.

            The subject invention will be illustrated
25 with synthetic fragments coding for IGF I and IGF II
joined to the leader and processing signals of yeast
α-factor. The yeast α-factor may be restricted with
HindIII and SalI. HindIII cleaves in the processing
signal of the α-factor precursor, cleaving 3' to the

30 second base in the coding strand of the glu codon,
while the HindIII recognition sequence completes the
glu codon, encodes for ala and provides the first 5'
base of the amino-terminal trp codon of mature α-factor.
With reference to the direction of transcription of the

35 α-factor gene, the SalI site is located upstream of the
transcriptional terminator.

The synthetic genes coding for IGF will have nucleotide sequences based on the known amino acid sequences of the IGF I and IGF II polypeptides. Preferably, the synthetic sequences will employ codons

5   which are preferentially utilized by the yeast host, e.g., based on the frequency with which the codons are found in the genes coding for the yeast glycolytic enzymes. Conveniently, the synthetic sequence will include cohesive ends rather than blunt ends for inser-

10   tion into a restriction site in a cloning vehicle. Furthermore, restriction sites will be designed into the synthetic sequences using silent mutations in order to generate fragments which may be annealed into sequences capable of producing IGF I/IGF II hybrid

15   peptide molecules.

In the examples, the synthetic fragments are provided with cohesive ends for EcoRI and inserted into the EcoRI site in pBR328. Usually, the synthetic se- quence will include additional restriction sites proxi-

20   mal to each end of the polypeptide coding region. Such interior restriction sites are selected to provide precise excision of the coding region from the cloning vehicle and for joining to adaptors so that the final DNA construct, including the leader and processing

25   signals and coding region, are in proper reading frame, and in proper juxtaposition to a transcription termina- tor. Preferably, the restriction sites will have the recognition sequence offset from the cleavage site, where cleavage is directed proximal to the coding

30   region and the recognition site is lost. This allows cleavage precisely at each end of the coding region regardless of the nucleotide sequence. BgaI sites are provided in the examples.

In preparing the synthetic gene, overlapping

35   single stranded DNA (ssDNA) fragments are prepared by conventional techniques. Such ssDNA fragments will usually be from about 10 to 40 bases in length.

Although considerably longer fragments may be employed, the synthetic yield decreases and it becomes more difficult to assure that the proper sequence has not been inadvertently degraded or altered. After the

5   ssDNA fragments have been synthesized, they are joined under annealing conditions with complementary base pairing assuring the proper order. The ends of the fragments are then ligated, and the resulting synthetic DNA fragment cloned and amplified, usually in a bacterial

10   host such as E. coli. As previously indicated, the synthetic structural gene may be provided with cohesive ends complementary to a suitable restriction site in the cloning vehicle of interest and internal recognition sites which allow for precise excision of the coding

15   region. After cloning and amplification of the synthetic sequences, usable quantities of the sequences may be excised, usually at the internal restriction sites on either end of the IGF coding region.

Conveniently, the promoter which is employed

20   may be the promoter associated with the leader and processing sequence. In this manner, a 5'-portable element, which contains both the promoter and the leader sequence in proper spatial relationship for efficient transcription, may be provided. By further

25   including a transcriptional terminator, a "cassette" consisting of promoter/leader - restriction site(s) - terminator is created, where the IGF coding region may be inserted with the aid of adaptors. Usually, such cassettes may be provided by isolating a DNA fragment

30   which includes an intact gene from a yeast host and the upstream and downstream transcriptional regulatory sequences of the gene, where the gene expresses a polypeptide which is secreted by the host.

Alternatively, one may replace the naturally

35   occurring yeast promoter by other promoters which allow for transcriptional regulation. This will require sequencing and/or restriction mapping of the region

upstream from the leader sequence to provide for intro-
duction of a different promoter. In some instances,
it may be desirable to retain the naturally occurring
yeast promoter and provide a second promoter in tandem,
5   either upstream or downstream from the naturally
occurring yeast promoter.

A wide variety of promoters are available or
can be obtained from yeast genes. Promoters of partic-
ular interest include those promoters involved with
10   enzymes in the glycolytic pathway, such as promoters
for alcohol dehydrogenase, glyceraldehyde-3-phosphate
dehydrogenase, pyruvate kinase, triose phosphate
isomerase, phosphoglucoisomerase, phosphofructokinase,
etc. By employing these promoters with regulatory
15   sequences, such as enhancers, operators, etc., and
using a host having an intact regulatory system, one
can regulate the expression of the processed "pre"-IGF.
Thus, various small organic molecules, e.g. glucose,
may be employed for the regulation of production of the
20   desired polypeptide.

One may also employ temperature-sensitive
regulatory mutants which allow for modulation of
transcription by varying the temperature. Thus, by
growing the cells at the non-permissive temperature,
25   one can grow the cells to high density, before changing
the temperature in order to provide for expression of
the "pre"-polypeptides for IGF I and IGF II.

Other capabilities may also be introduced
into the construct. For example, some genes provide
30   for amplification, where upon stress to the host, not
only is the gene which responds to the stress amplified,
but also flanking regions. By placing such a gene
upstream from the promoter, coding region and the other
regulatory signals providing transcriptional control of
35   the "pre"-polypeptide, and stressing the yeast host,
plasmids may be obtained which have a plurality of
repeating sequences, which sequences include the

"pre"-polypeptide gene with its regulatory sequences. Illustrative genes include metallothioneins and dihydro-folate reductase.

The construct may include in addition to the
5 leader sequence fragment, other DNA homologous to the host genome. If it is desired that there be integration of the IGF gene into the chromosome(s), integration can be enhanced by providing for sequences flanking the IGF gene construct which are homologous to host chromosomal
10 DNA.

The replication system which is employed will be recognized by the yeast host. Therefore, it is desirable that the replication system be native to the yeast host. A number of yeast vectors are reported by
15 Botstein et al., Gene (1979) 8:17-24. Of particular interest are the YEp plasmids, which contain the 2μm plasmid replication system. These plasmids are stably maintained at multiple copy number. Alternatively or in addition, one may use a combination of ARS1 and
20 CEN4, to provide for stable maintenance.

After each manipulation, as appropriate, the construct may be cloned so that the desired construct is obtained pure and in sufficient amount for further manipulation. Desirably, a shuttle vector (i.e.,
25 containing both a yeast and bacterial origin of repli-cation) may be employed so that cloning can be performed in prokaryotes, particularly E. coli.

The plasmids may be introduced into the yeast host by any convenient means, employing yeast host
30 cells or spheroplasts and using calcium precipitated DNA for transformation or liposomes or other conventional techniques. The modified hosts may be selected in accordance with the genetic markers which are usually provided in a vector used to construct the expression
35 plasmid. An auxotrophic host may be employed, where the plasmid has a gene which complements the host and provides prototrophy. Alternatively, resistance to an

appropriate biocide, e.g. antibiotic, heavy metal, toxin, or the like, may be included as a marker in the plasmid. Selection may then be achieved by employing a nutrient medium which stresses the parent cells, so as

5 to select for the cells containing the plasmid. The plasmid containing cells may then be grown in an appropriate nutrient medium, and the desired secreted polypeptide isolated in accordance with conventional techniques. The polypeptide may be purified by chroma-

10 tography, filtration, extraction, etc. Since the polypeptide will be present in mature form in the nutrient medium, one can cycle the nutrient medium, continuously removing the desired polypeptide.

The following examples are offered by way of

15 illustration and not by way of limitation.

EXPERIMENTAL

Nucleotide sequences for human insulin-like growth factors I and II (IGF I and II) comprising preferred yeast codons were devised based on the amino

20 acid sequences reported in Rinderknecht and Humbel (1978) J.Biol.Chem. 253:2769-2776 and Rinderknecht and Humbel (1978) FEBS Letters 89:283-286, respectively. The sequences (with the coding strands shown 5' to 3') are as follows:
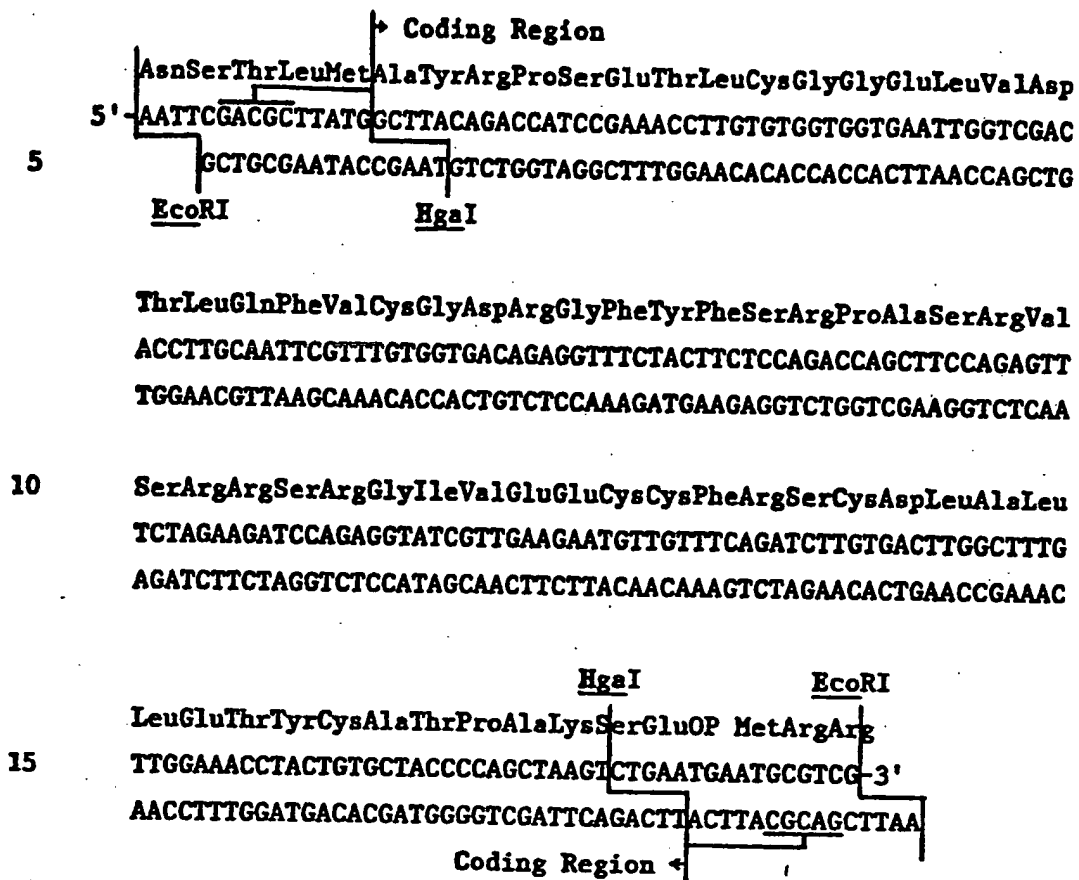
IGF I

```
                                 → Coding Region
     AsnSerThrLeuMetGlyProGluThrLeuCysGlyAlaGluLeuValAspAlaLeuGln
5'- AATTCGACGCTTATGGGTCCAGAAACCTTGTGTGGTGCTGAATTGGTCGATGCTTTGCAA
5       GCTGCGAATACCCAGGTCTTTGGAACACACCACGACTTAACCAGCTACGAAACGTT
        EcoRI                   HgaI
```

```
     PheValCysGlyAspArgGlyPheTyrPheAsnLysProThrGlyTyrGlySerSerSer
     TTCGTTTGTGGTGACAGAGGTTTCTACTTCAACAAGCCAACCGGTTACGGTTCTTCTTCT
     AAGCAAACACCACTGTCTCCAAAGATGAAGTTGTTCGGTTGGCCAATGCCAAGAAGAAGA
```

```
10   ArgArgAlaProGlnThrGlyIleValAspGluCysCysPheArgSerCysAspLeuArg
     AGAAGAGCTCCACAAACCGGTATCGTTGACGAATGTTGTTTCAGATCTTGTGACTTGAGA
     TCTTCTCGAGGTGTTTGGCCATAGCAACTGCTTACAACAAAGTCTAGAACACTGAACTCT
```

```
                                          HgaI              EcoRI
     ArgLeuGluMetTryCysAlaProLeuLysProAlaLysSerAlaOP MetArgArg
15   AGATTGGAAATGTACTGTGCTCCATTGAAGCCAGCTAAGTCTGCTTGAATGCGTCG-3'
     TCTAACCTTTACATGACACGAGGTAACTTCGGTCGATTCAGACGAACTTACGCAGCTTAA
                             Coding Region ←
```

IGF II

→ Coding Region

```
          AsnSerThrLeuMet AlaTyrArgProSerGluThrLeuCysGlyGlyGluLeuValAsp
     5'--AATTCGACGCTTATGGCTTACAGACCATCCGAAACCTTGTGTGGTGGTGAATTGGTCGAC
5        GCTGCGAATACCGAATGTCTGGTAGGCTTTGGAACACACCACCACTTAACCAGCTG
     EcoRI              HgaI
```

```
          ThrLeuGlnPheValCysGlyAspArgGlyPheTyrPheSerArgProAlaSerArgVal
          ACCTTGCAATTCGTTTGTGGTGACAGAGGTTTCTACTTCTCCAGACCAGCTTCCAGAGTT
          TGGAACGTTAAGCAAACACCACTGTCTCCAAAGATGAAGAGGTCTGGTCGAAGGTCTCAA
```

```
10        SerArgArgSerArgGlyIleValGluGluCysCysPheArgSerCysAspLeuAlaLeu
          TCTAGAAGATCCAGAGGTATCGTTGAAGAATGTTGTTTCAGATCTTGTGACTTGGCTTTG
          AGATCTTCTAGGTCTCCATAGCAACTTCTTACAACAAAGTCTAGAACACTGAACCGAAAC
```

```
                                        HgaI           EcoRI
          LeuGluThrTyrCysAlaThrProAlaLysSerGluOP MetArgArg
15        TTGGAAACCTACTGTGCTACCCCAGCTAAGTCTGAATGAATGCGTCG-3'
          AACCTTTGGATGACACGATGGGGTCGATTCAGACTTACTTACGCAGCTTAA
                         Coding Region ←
```

The sequences are provided with EcoRI cohesive ends at both ends. Coding for IGF I begins at base 16 of the coding strand and ends at base 225. HgaI restriction sites are located at each end of the IGF I coding region. The HgaI recognition sites (5'-GACGC-3') lie outside of the IGF I coding region, i.e. between the end of the synthetic sequence and the HgaI cleavage site. The IGF II synthetic sequence is similarly constructed with the coding strand beginning at base 16 and terminating at base 219.

A synthetic DNA fragment for IGF I having the sequence just described for IGF I was prepared by synthesizing 20 overlapping ssDNA segments using the phosphoramidite method (see co-pending application,

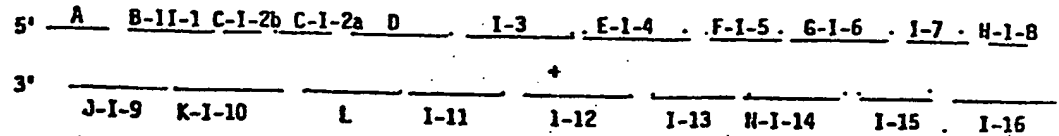Ser. No. 457,412, filed January 12, 83). The ssDNA
sequences were as follows:

	Designation	Sequence
	A	AATTCGACGCTTATGG
5	B-I-1	GTCCAGAAACCTTGTGTGGT
	C-I-2b	GCTGAATTGGTC
	C-I-2a	GATGCTTTGCAATTCGT
	D	TTGTGGTGACAGAGGTTTCTACTTC
	I-3	AACAAGCCAACCGGTTACGGTTCTTCTTC
10	E-I-4	TAGAAGAGCTCCACAAACCGGTATCGTT
	F-I-5	GACGAATGTTGTTTCAGATCTTGTGACTTG
	G-I-6	AGAAGATTGGAAATGTACTGTGCT
	I-7	CCATTGAAGCCAGCTAAGTCT
	H-I-8	GCTTGAATGCGTCG
15	J-I-9	GTTTCTGGACCCATAAGCGTCG
	K-I-10	AAAGCATCGACCAATTCAGCACCACACAAG
	L	CTCTGTCACCACAAACGAATTGC
	M-I-11	AACCGGTTGGCTTGTTGAAGTAGAAAC
	I-12	TGGAGCTCTTCTAGAAGAAGAACCGT
20	I-13	GAAACAACATTCGTCAACGATACCGGTTTG
	N-I-14	CCAATCTTCTCAAGTCACAAGATCT
	I-15	GGCTTCAATGGAGCACAGTACATTT
	I-16	AATTCGACGCATTCAAGCAGACTTAGCT

The ssDNA segments were joined as follows:
25 50 pmoles of each segment except A and I-16 were
5'-phosphorylated with T4 polynucleotide kinase.
50pmoles all segments were then annealed in a single
step as an 18µl pool by cooling from 95° to 25° over
1.5 hrs. Ligation was achieved in a reaction volume of
30 30µl containing 1mM ATP, 10mM DTT, 100mM tris-HCl, pH
7.8, 10mM MgCl$_2$, 1µg/ml spermidine and T4 ligase.
The appropriate double-stranded fragment resulting from
the order and pairing of fragments shown in Figure 1
was purified on a 7% native polyacrylamide electrophore-
35 sis gel.

0123228

Figure 1

IGF I ANNEALING AND LIGATION SCHEME

```
5'  A   B-II-1 C-I-2b C-I-2a  D      I-3    . E-1-4 . F-I-5 . 6-I-6 . I-7 . H-1-8
    ___ _____ _____ _____ ___    ___    _____   _____   _____   ___   _____
                                      +
3'  ___ _____ _____ _____   ___    ___    _____   _____   ___    ___    _____
    J-I-9 K-I-10      L      I-11    1-12    I-13  N-I-14    I-15 . I-16
```

A DNA sequence for IGF II was similarly
synthesized. The following additional ssDNA fragments
were prepared.

	Designation	Sequence
5	B-II-1	CTTACAGACCATCCGAAACCTTGTGTGGT
	C-II-2	GGTGAATTGGTCGACACCTTGCAATTCGT
	II-3.	TCCAGACCAGCTTCCAGAGTTTCT
	E-II-4	AGAAGATCCAGAGGTATCGTT ·
	F-II-5	GAAGAATGTTGTTTCAGATCTTGTGACTTG
10	G-II-6	GCTTTGTTGGAAACCTACTGTGCT
	H-II-7	ACCCCAGCTAAGTCTGAATGAATGCGTCG
	J-II-8	GTTCGGATGGTCTGTAAGCCATAAGCGTCG
	K-II-9	AAGGTGTCGACCAATTCACCACCACACAAG
	M-II-10	AAGCTGGTCTGGAGAAGTAGAAAC
15	II-11	CTGGATCTTCTAGAAACTCTGG
	II-12	GAAACAACATTCTTCAACCOATACCT
	N-II-13	TTCCAACAAAGCCAAGTCACAAGATCT
	II-14	AGACTTAGCTGGGGTAGCACAGTAGGT
	II-15	AATTCGACGCATTCATTC ·

20   200 pmoles of these ssDNA fragments and
fragments A, I and L were joined in a similar manner as

above, wherein A and II-15 were not phosphorylated
resulting in the following ordering and pairing of
segments.

Figure 2

IGF II ANNEALING AND LIGATION SCHEME

```
 A   B-II-1  C-II-2    D     II-3   E-II-4  F-II-5  G-II-6   H-II-7
                                                                    3'
                                      +
   J-II-8  K-II-9   L   M-II-10  II-11  II-12  N-II-13   II-14   II-15 5'
```
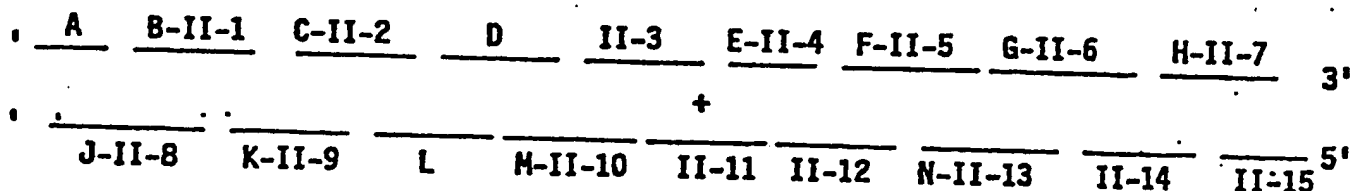
The synthetic DNA sequences were inserted
5   into the EcoRI site of pBR328 to produce plasmids
p328IGF I and p328IGF II. After cloning, the IGF
coding strands were excised using HgaI.

Synthetic oligonucleotide adaptors were then
ligated to the HgaI restriction fragments. For IGF I,
10 the adaptors had the following sequences:

        a)   5'-AGCTGAAGCT-3'
                 3'-CTTCGACCAGG-5'

        b)   5'-CTGCTTGATAAG-3'
                 3'-ACTATTCAGCT-5'

15 With orientation with respect to the coding strand, the
3'-end of the first adaptor, a), is complementary to
the 5' HgaI cleavage site on the IGF I synthetic
sequence, while the 5'-end of the first adaptor provides
a HindIII cohesive end. The second adaptor, b), is

complementary at its 5'-end to the HgaI cleavage site
at the 3'-end of the IGF I sequence, while the 3'-end
of the adaptor provides a SalI cohesive end.

For IGF II, the adaptors had the following
5 sequences:

c)   5'-AGCTGAAGCT-3'
           3'-CTTCGACGAAT-5'

d)   5'-CTGAATGATAAG-3'
           3'-ACTATTCAGCT-5'

10 With orientation as above the first adaptor, c), is
complementary at its 3'-end to the 5'- HgaI cleavage
site in the IGF II synthetic sequence, while the 5'-end
provides a HindIII cohesive end. The second adaptor,
d), is complementary at its 5'-end to the second HgaI
15 cleavage site in the IGF II synthetic sequence and
provides a SalI cohesive end at its 3'-end.

The synthetic fragments and adaptors joined
thereto were purified by preparative gel electrophoresis
and ligated to 100ng of pAB113 which had been previously
20 digested to completion with endonucleases HindIII and
SalI. pAB113 was derived from pAB112 by deletion of
the three 63 bp HindIII fragments. pAB112 is a plasmid
containing a 1.8kb EcoRI fragment with the yeast
α-factor gene cloned in the EcoRI site of pBR322 in
25 which the HindIII and SalI sites had been deleted.
pAB112 was derived from plasmid pAB101 which contains
the yeast α-factor gene as a partial Sau3A fragment
cloned in the BamHI site of plasmid YEp24. pAB101 was
obtained by screening a yeast genomic library cloned in
30 YEp24 using an enzymatically P^{32} radiolabeled synthetic
oligonucleotide probe homologous to the published
α-factor coding region (Kurjan and Herskowitz, Abstracts
1981 Cold Spring Harbor meeting on the Molecular
Biology of Yeasts, page 242).

35   The resulting mixtures were used to transform
E. coli HB101 cells, and plasmids pAB113-IGF-I and

pAB113-IGF-II were obtained for IGF I and IGF II,
respectively.

Plasmids pAB113-IGF-I and pAB113-IGF-II (5μg
each) having the IGF I and IGF II structural genes,
5   respectively, were digested to completion with EcoRI
and the resulting fragments were ligated to an excess
of EcoRI-BamHI adaptors and digested with BamHI. The
resultant 1.8kb BamHI fragments were isolated by
preparative gel electrophoresis and approximately 100ng
10   of each fragment was ligated to 100ng of pCl/1, which
had been previously digested to completion with BamHI·
and treated with alkaline phosphatase.

Plasmid pCl/1 is a derivative of pJDB219,
Beggs, Nature (1978) 275:104, in which the region
15   corresponding to bacterial plasmid pMB9 in pJDB219 has
been replaced by pBR322 in pCl/1. Each ligation
mixture was used to transform E. coli HB101 cells.
Transformants were selected by ampicillin resistance
and their plasmids analyzed by restriction endonuclease
20   digestion. DNA from one selected clone for each
structural gene, IGF I or IGF II, (pYIGF-I-10/1) or
(pYIGF-II-10/1), respectively, was prepared and used to
transform yeast AB103 cells. Transformants were
selected by their Leu$^+$ phenotype.

25   Two cultures (5 and 9 liters) of yeast strain
AB103 (α,pep 4-3, leu 2-3, leu 2-112, ura 3-52, his
4-580) transformed with plasmid (pYIGF-I-10/1) were
grown at 30°C in -leu medium to saturation (optical
density at 650nm of 5) and left shaking at 30°C for an
30   additional 12 hr period. Cell supernatants were
collected from each culture by centrifugation and the
IGF I concentrated by absorption on an ion exchange
resin (Biorex-70 available from Bio-Rad Laboratories,
Inc., Richmond, California). After elution with 10mM
35   HCl in 80% ethanol, the IGF I fractions (0.4ml and 3ml,
respectively) were assayed for total protein concentra-
tion and IGF I concentration. The protein assay was

the Coomassie Blue assay available from Bio-Rad Labora-
tories, Inc., Richmond, California. The IGF I assay
was a conventional competitive radioimmunoassay employing
radiolabeled IGF I. The results were as follows:

5

Trial No.	Volume	Volume After Concentration	Total Protein	IGF-I
1	5 liters	0.4ml	5.0mg/ml	3.75mg/ml
2	9 liters	3.0ml	2.9mg/ml	3.00mg/ml

A bioassay of IGF I, based on the synergystic effect of
10 the peptide to promote the response of pigeon crop sac
to prolactin (Anderson et al. (1983) in Somatomedins/
Insulin-Like Growth Factors, Spencer, E.M., ed., Walter
deGruter, Berlin) reveals that the IGF I product of
these preparations has activity equivalent to authentic
15 IGF I isolated from human serum.

Cultures of AB103(pYIGF-II-10/1) grown
similarly to the above and assayed using a human
placental membrane radio receptor assay (Spencer et al.
(1979) Act. Endocrinol. 91:36-48) for IGF II revealed
20 3.9 units/ml where normal human serum possesses 1
unit/ml.

In accordance with the subject invention,
novel constructs are provided which may be inserted
into vectors to provide for expression of human insulin-
25 like growth factor I to provide processing and secretion.
Thus, one can obtain a polypeptide having the identical
sequence to the naturally occurring human insulin-like
growth factor I. By virtue of providing for secretion,
greatly enhanced yields can be obtained based on cell
30 population and subsequent preparative operations and
purification are simplified.

0123228

Although the foregoing invention has been
described in some detail by way of illustration and
example for purposes of clarity of understanding, it
will be obvious that certain changes and modifications

5   may be practiced within the scope of the appended
claims.

WHAT IS CLAIMED IS:

    1.   A method for efficient production of a human insulin-like growth factor (IGF) in a suitable host, said method comprising growing host cells contain-
5 ing a functional DNA construct having the gene encoding said IGF joined in proper reading frame with secretory leader and processing signal sequences recognized by said host to form a structural gene downstream from and under the transcriptional regulatory control of a
10 transcriptional initiation region in a vector compatible with said host, and isolating secreted human IGF.

    2.   A method as in claim 1, wherein the IGF gene is a synthetic gene having codons preferentially utilized by the host.

15   3.   A method as in claim 1, wherein the human IGF gene is a synthetic IGF I gene having the following nucleotide sequence:

```
          GlyProGluThrLeuCysGlyAlaGluLeuValAspAlaLeuGln
          5'-GGTCCAGAAACCTTGTGTGGTGCTGAATTGGTCGATGCTTTGCAA
20              CCAGGTCTTTGGAACACACCACGACTTAACCAGCTACGAAACGTT


PheValCysGlyAspArgGlyPheTyrPheAsnLysProThrGlyTyrGlySerSerSer
TTCGTTTGTGGTGACAGAGGTTTCTACTTCAACAAGCCAACCGGTTACGGTTCTTCTTCT
AAGCAAACACCACTGTCTCCAAAGATGAAGTTGTTCGGTTGGCCAATGCCAAGAAGAAGA


ArgArgAlaProGlnThrGlyIleValAspGluCysCysPheArgSerCysAspLeuArg
25 AGAAGAGCTCCACAAACCGGTATCGTTGACGAATGTTGTTTCAGATCTTGTGACTTGAGA
TCTTCTCGAGGTGTTTGGCCATAGCAACTGCTTACAACAAAGTCTAGAACACTGAACTCT


ArgLeuGluMetTyrCysAlaProLeuLysProAlaLysSerAla
AGATTGGAAATGTACTGTGCTCCATTGAAGCCAGCTAAGTCTGCT-3'
TCTAACCTTTACATGACACGAGGTAACTTCGGTCGATTCAGACGA
```

4. A method as in claim 1, wherein the human IGF gene is a synthetic IGF II gene having the following nucleotide sequence:

AlaTyrArgProSerGluThrLeuCysGlyGlyGluLeuValAsp

5 5'-GCTTACAGACCATCCGAAACCTTGTGTGGTGGTGAATTGGTCGAC

CGAATGTCTGGTAGGCTTTGGAACACACCACCACTTAACCAGCTG

ThrLeuGlnPheValCysGlyAspArgGlyPheTyrPheSerArgProAlaSerArgVal

ACCTTGCAATTCGTTTGTGGTGACAGAGGTTTCTACTTCTCCAGACCAGCTTCCAGAGTT

TGGAACGTTAAGCAAACACCACTGTCTCCAAAGATGAAGAGGTCTGGTCGAAGGTCTCAA

10 SerArgArgSerArgGlyIleValGluGluCysCysPheArgSerCysAspLeuAlaLeu

TCTAGAAGATCCAGAGGTATCGTTGAAGAATGTTGTTTCAGATCTTGTGACTTGGCTTTG

AGATCTTCTAGGTCTCCATAGCAACTTCTTACAACAAAGTCTAGAACACTGAACCGAAAC

LeuGluThrTyrCysAlaThrProAlaLysSerGlu

TTGGAAACCTACTGTGCTACCCCAGCTAAGTCTGAA-3'

15 AACCTTTGGATGACACGATGGGGTCGATTCAGACTT

5. A method as in claim 1, wherein the host is a yeast.

6. A method as in claim 1, wherein the host is a yeast and the vehicle is a derivative of the 2μm

20 plasmid.

7. A method for expressing human insulin-like growth factor (IGF) in yeast resulting in secretion of said IGF promoted by yeast secretion and processing signals, said method comprising:

25   preparing a first DNA fragment containing a first DNA sequence encoding for IGF and having the 5'-terminus of the coding strand at or downstream from the first base of said DNA sequence;

preparing a second DNA fragment containing a

30 second DNA sequence encoding for yeast-recognizable

secretion and processing signals and having the 3'-
terminus of the coding strand at or upstream from the
terminal base of said processing signal, wherein at
least one of said first and second fragments has its
5   terminus internal to the coding sequence resulting in
missing base pairs;

joining said first and second DNA fragments
by means of an adaptor encoding for the missing base
pairs of said first and second DNA fragments, wherein
10   said first and second DNA fragments are in the same
reading frame;

providing a termination codon downstream of
and adjacent to the first fragment to provide a "pre"-IGF
gene; and

15   cloning said "pre"-IGF gene in an expression
vector in yeast, wherein said "pre"-IGF is secreted and
processed.

8.   A method according to claim 7, wherein
said IGF is IGF I.

20   9.   A method according to claim 7, wherein
said IGF is IGF II.

10.   A method according to claim 7, wherein
said expression vector includes a replication system
recognized by bacteria.

25   11.   A method according to claim 7, wherein
said yeast replication system comprises the 2μm plasmid
or portion thereof.

12.   A DNA construct comprising a coding
sequence coding for human insulin-like growth factor
30   (IGF) under transcriptional regulation of regulatory
signals utilized by yeast.

13. A DNA construct according to claim 12, wherein said IGF coding sequence codes for IGF I.

14. A DNA construct according to claim 12, wherein said IGF coding sequence codes for IGF II.

5 15. A DNA construct according to claim 12 including a replication system recognized by yeast.

16. A DNA construct according to claim 12, where at least a portion of said coding sequence has codons preferentially utilized by yeast and different

10 from human codons.

17. A DNA construct according to claim 12, wherein said construct includes a replication system utilized by bacteria.

18. A DNA construct according to claim 15

15 wherein said yeast replication system comprises the 2μm plasmid or portion thereof.

19. A DNA construct according to claim 12, including the sequence:

GlyProGluThrLeuCysGlyAlaGluLeuValAspAlaLeuGln
5'-GGTCCAGAAACCTTGTGTGGTGCTGAATTGGTCGATGCTTTGCAA
   CCAGGTCTTTGGAACACACCACGACTTAACCAGCTACGAAACGTT

PheValCysGlyAspArgGlyPheTyrPheAsnLysProThrGlyTyrGlySerSerSer
5 TTCGTTTGTGGTGACAGAGGTTTCTACTTCAACAAGCCAACCGGTTACGGTTCTTCTTCT
   AAGCAAACACCACTGTCTCCAAAGATGAAGTTGTTCGGTTGGCCAATGCCAAGAAGAAGA

ArgArgAlaProGlnThrGlyIleValAspGluCysCysPheArgSerCysAspLeuArg
AGAAGAGCTCCACAAACCGGTATCGTTGACGAATGTTGTTTCAGATCTTGTGACTTGAGA
TCTTCTCGAGGTGTTTGGCCATAGCAACTGCTTACAACAAAGTCTAGAACACTGAACTCT

10 ArgLeuGluMetTyrCysAlaProLeuLysProAlaLysSerAla
    AGATTGGAAATGTACTGTGCTCCATTGAAGCCAGCTAAGTCTGCT-3'
    TCTAACCTTTACATGACACGAGGTAACTTCGGTCGATTCAGACGA—

    20.   A DNA construct according to claim 12,
including the sequence:

15   AlaTyrArgProSerGluThrLeuCysGlyGlyGluLeuValAsp
           5'-GCTTACAGACCATCCGAAACCTTGTGTGGTGGTGAATTGGTCGAC
             CGAATGTCTGGTAGGCTTTGGAACACACCACCACTTAACCAGCTG

ThrLeuGlnPheValCysGlyAspArgGlyPheTyrPheSerArgProAlaSerArgVal
ACCTTGCAATTCGTTTGTGGTGACAGAGGTTTCTACTTCTCCAGACCAGCTTCCAGAGTT
20 TGGAACGTTAAGCAAACACCACTGTCTCCAAAGATGAAGAGGTCTGGTCGAAGGTCTCAA

SerArgArgSerArgGlyIleValGluGluCysCysPheArgSerCysAspLeuAlaLeu
TCTAGAAGATCCAGAGGTATCGTTGAAGAATGTTGTTTCAGATCTTGTGACTTGGCTTTG
AGATCTTCTAGGTCTCCATAGCAACTTCTTACAACAAAGTCTAGAACACTGAACCGAAAC

LeuGluThrTyrCysAlaThrProAlaLysSerGlu
25 TTGGAAACCTACTGTGCTACCCCAGCTAAGTCTGAA-3'
   AACCTTTGGATGACACGATGGGGTCGATTCAGACTT

21.   A DNA construct comprising a sequence coding for human insulin-like growth factor joined to transcriptional and translational regulatory sequences for expression of a cellular host in culture.